

In the Specification

Applicants request that the specification be amended by replacing the corresponding numbered paragraphs with the following numbered paragraphs.

[0079] FIG. 12 is a flowchart of the creation of the difference file of step 762 of FIG. 11 of an embodiment of the present invention. At step 810 a delimiter that is used to partition the edited and unedited replica into segments is selected. The delimiter is picked for each document, e.g., unedited replica, by a statistical process to try to avoid too many very short or very long segments. The same delimiter is used for both documents in a pair, but for different pairs of documents the delimiter may differ. At step 812 the unedited replica is partitioned into segments, using the delimiter. Additional segments are also created for contiguous runs of matching bytes, for example, a series of 7 bytes each containing the value zero would be a segment. At step 814 the unedited replica segments are sorted into a binary tree.

[0080] The binary tree in one embodiment is based on a "less than" function, i.e., one segment is determined to be "less than" another segment based upon length and content. This function conforms to the mathematical '<' operator. For example if $A < B$ and $B < C$, then $A < C$, where A, B, and C are segments. Once all the segments in the edited unedited replica are built into the binary tree, then a given segment in the edited replica can be searched for a match in the binary tree in order $\log_{2.5} N$ time (where N is the number of segments in the binary tree). In one aspect this binary tree has a

dictionary like ordering. So that if the documents consisted only of plain text, a segment containing the text `bee` would appear after `be` but before `beg`.

[0081] At step 816 the edited replica is partitioned into segments similar to step 812. Next each segment in the edited replica is compared with the segments in the binary tree. In one embodiment each segment in the edited replica is compared. In an alternative embodiment some segments may be skipped as it may not be necessary to compare them to segments in the binary tree. The end result is to get a series of copy packets and difference packets. The copy and difference packets are part of the difference file and will be used by the manager application to reconstruct the edited replica from the unedited replica. A copy packet indicates that a copy of a portion of the unedited replica should be inserted at the current position in the edited replica, that is being reconstructed. The copy packet specifies the position and size (in bytes) of the portion of the unedited replica to be inserted. A difference packet indicates that a stream of bytes from the edited replica, which were not found in the unedited replica, should be inserted at the current position in the edited replica that is being reconstructed. The difference packet contains the size of the stream of bytes to be inserted as well as the stream of bytes itself.

[0091] The difference packet would be the stream of bytes between the two copy packets or "hj; kk;" or [hj; kk]. Thus the final difference file is: [0,5] [hj; kk][8,6].